

教育数据挖掘研究现状及趋势

丁波 孙力

(江南大学 人文学院, 江苏 无锡 214000)

摘要: 本文对教育数据挖掘 (Educational Data Mining, EDM) 的兴起、发展、未来研究趋势等进行了最新的概述, 旨在为各类群体进行教育数据挖掘研究或实践提供实时、有价值的参考。本文首先对 EDM 研究背景进行了阐述, 然后重点介绍了影响 EDM 应用的三要素: 教育数据挖掘目标、教育数据来源、教育数据挖掘过程, 最后分析了 EDM 主要应用及未来研究趋势。

关键词: 教育数据挖掘; 影响要素; 应用; 研究趋势

中图分类号: G4 **文献标识码:** A **文章编号:** 2096-0069 (2015) 06-0013-04

一、教育数据挖掘 (EDM) 研究背景及现状

(一) 教育数据挖掘简介

随着在线教育的成熟、网络教育资源的增加、数据应用的普及, 各类教育系统中产生了大量特殊类型的数据。通过对这些教育数据的分析, 教育工作者可以对学生及其如何学习进行更深入的探究。^[1]

教育数据挖掘是一个跨学科领域, 不仅涉及信息检索、可视化数据分析、数据驱动、社会网络分析, 还涉及教育心理学、认知心理学、心理统计学等。

近年来, EDM 已成为一个独立的研究领域, 其目的是分析教育系统中产生的独特类型的数据来解决教育研究问题。EDM 往往也涉及算法的开发、研究及应用,

利用算法构建并集成出教育数据模型, 从而通过模型分析出海量数据背后隐藏的意义。因此, 教育数据挖掘可以定义为, 应用数据挖掘技术获取来自教育系统的特殊类型数据, 提取有意义的信息并分析出有价值的结果, 这些分析结果可供学习者、教育工作者、教育软件开发、教育管理者、教育研究者等对象利用。^[2]

(二) 研究背景及现状

教育数据挖掘起源于对智能教学系统 (ITS)、人工智能教育 (AIED)、用户模型 (UM)、技术增强学习 (TEL)、自适应智能教育超媒体 (AIEH) 等的研究。^[3] 直至 2008 年, “教育数据挖掘国际会议” 独立召开, 不再依附于计算机应用学术会议。

目前, 关于 EDM 的书籍只出版了寥寥数本。

收稿日期: 2015-09-30

作者简介: 丁波 (1989—), 女, 江苏泰州人, 江南大学人文学院教育技术学硕士研究生, 研究方向为网络教育系统的构建及开发; 孙力 (1966—), 男, 江苏无锡人, 博士, 江南大学人文学院教授, 继续教育与网络教育学院副院长, 研究方向为网络教育系统的构建及开发。

国外出版了 *Data Mining in E-Learning* (《教育数据挖掘》) 和 *Handbook of Educational Data Mining* (《教育数据挖掘手册》), 国内出版了《教育数据挖掘: 方法与应用》。近年来, 关于 EDM 的文章在一些具有广泛国际影响、与 EDM 主题相关的著名杂志上大量发表。如 JEDM (《教育数据挖掘》杂志, *Journal of Educational Data Mining*), JAIED (《教育人工智能》杂志, *Journal of Artificial Intelligence in Education*), JLS (《学习科学》杂志, *Journal of the Learning Sciences*), CAE (《计算机与教育》杂志, *Computer and Education*), KDE (《知识与数据工程》杂志, *IEEE Transactions on Knowledge and Data Engineering*), 等等。

随着 EDM 的发展, 对该领域有重要贡献的学者也越来越多。

在谷歌学术和知网分别检索关键字“*educational data mining*”和“教育数据挖掘”, 获取从 2008 年到 2015 年来的论文数量, 并对论文数量进行对比, 可以看出谷歌学术和知网内的关于 EDM 文章的数量都在以指数方式增长, 这显示出人们对于 EDM 的高度关注。目前关于 EDM 的文章中有许多理论性及综述类论文, 不过涉及 EDM 具体应用的论文还较少。

二、教育数据挖掘影响要素分析

依据 EDM 研究教育问题的思路, 可以将教育数据挖掘目标、教育数据来源、教育数据挖掘过程作为影响 EDM 应用的三要素。

(一) 教育数据挖掘目标

EDM 涉及不同的对象群体。不同群体根据他们的任务、愿景和目标, 从不同的角度看待教育信息并使用数据挖掘技术。

对学习者的而言, EDM 的目标是实现个性化在线学习; 向学习者推荐学习活动、学习资源及学习任务从而进一步提高他们的学习效率; 提供有意义的学习经验; 提供自适应提示、推荐课程和相关讨论、书籍等。对于教育者而言, EDM 的目标是获得客观的教学反馈; 分析学生的学习和行为; 发现需要帮助的学生; 预测学生成绩; 对学习者的分组; 发现学习者的学习规律;

发现学习者最常犯的错误; 确定更有效的活动; 提供自适应内容和定制课程等。对于教育软件开发者和教育研究者而言, EDM 的目标是评估和维护课件; 提高学生的学习效率; 评价课程内容结构及其在学习过程中的有效性; 自动构建学习者模型和教育者模型; 对比数据挖掘技术以便为每个任务推荐最有用的 EDM 方法; 制定特定的教育数据挖掘工具等。对于管理者而言, EDM 的目标是: 开发最佳方式以组织人力物力资源以及教育资源; 更有效地利用资源; 提高教育计划、远程学习方法的有效性; 评估教师和课程; 设置参数以提高网站效率及适应性。

(二) 教育数据来源

EDM 支持的教育系统中存在着大量不同类型的数据, 这些数据不仅包括学习者与教育系统交互产生的数据, 也包括协作数据、管理数据、人口统计数据、情感数据等。这些数据是教育领域特定的数据, 具有一定的内在语义信息和层次结构, 有别于其他领域的数据。

国内外众多学者将 EDM 的数据来源根据教育环境分为传统学校数据和网络教育数据。

传统教育环境一般包括学前教育环境、中小学教育环境、高等教育环境等。这些教育环境, 主要通过课堂教学、课堂讨论、讲座等教育者与学习者面对面的交流中产生课堂教学数据。随着课堂教学中智能电子教具、课堂互动反馈系统的广泛使用, 课堂教学中师生进行教学活动时产生的丰富数据都会被自动收集保存^[4]。网络教育数据来自网络教育环境, 如智能与自适应网络教育系统、在线学习系统、在线教学系统等。通过这些系统, 能够获取学习者基本信息、登录信息、学习行为数据、网络课程信息, 以及一些日志文件、交互信息、过程性数据和管理数据。随着互联网技术、人工智能技术、大数据技术等技术的不断发展, 网络环境下获取的教育数据也越来越丰富。

(三) 教育数据挖掘过程

教育数据挖掘过程可以看作数据挖掘和知识发现过程, 其过程如图 1 所示。在这个过程中, 不仅仅是将数据转化为知识, 更要将挖掘出的知识作用于教育环境从而对教育环境进行改善, 提高学生的学习环境质量。

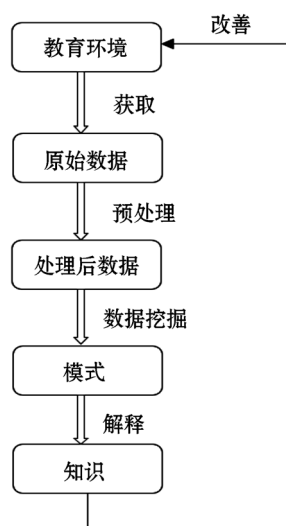


图1 教育数据挖掘过程

1. 数据获取

实施教育数据挖掘首先需从各类教育环境中收集不同类型的数据，这些数据包括管理数据、现场观测数据、问卷调查数据等。然而这些原始数据并不能直接使用，所以预处理是必要的。

2. 预处理

数据预处理既重要又复杂，教育数据结构、格式往往不符合数据挖掘要求或者不符合要解决问题的类型，因此必须先将数据转换为某种适当的形式，以解决特定的教育问题。数据预处理包括滤除冗余数据、填充丢失数据、统一数据格式等。

3. 数据挖掘

数据挖掘是整个 EDM 过程中非常重要的一环，在此过程中运用 EDM 技术，从数据中提取相应的知识。众多传统数据挖掘算法、技术已在教育领域广泛运用，如分类技术、聚类技术、关联分析技术等。然而，教育数据不同于其他领域的数据，所以需要一些特定的、区别于传统数据挖掘的方法来解决教育问题。

4. 数据解释

数据解释是最后一步也是非常重要的一步，利用发现的知识制定决策，改善教育环境，并将结果直观地显示给用户。这里可以利用可视化技术显示分析结果^[5]。

三、教育数据挖掘应用研究

EDM 已在教育领域中广泛应用。如贝克提出的四个

EDM 关键应用^[6]，卡斯特罗（Castro）提出的 EDM 主要应用^[7]。然而还有更多新型的 EDM 应用趋势，下面简要描述这些最新 EDM 应用及其涉及的 EDM 技术。

（一）给教师提供反馈

为教师提供反馈的目的是帮助教师做决策。关联规则是最常用的，利用关联规则进行教育数据挖掘可以揭示出变量间的关系，并根据不同的意义程度将它们以规则的形式呈现出来。另外，统计相关分析、模糊聚类分析、灰色关联分析、K-means 聚类分析和模糊关联分析这些数据挖掘技术已被应用于支持形成性评价^[8]，从而帮助教师理解影响学习者表现的主要影响因素，帮助教师开发更合理的试题、活动等。

（二）预测学生表现

预测学生的表现是 EDM 的一个典型应用，该应用中常用的 EDM 技术有神经网络技术、贝叶斯网络技术、回归相关分析等。预测学生表现的目的是通过评估一个变量未来的可能值来判断学生的表现，可以用来预测学生期末成绩，预测学生可能会犯的错误，预测学生正确回答问题的概率，预测学生可能进入的大学，等等。

（三）指导学生分组

对学生进行分组的目的是根据学生个人特征及定制内容创建不同的学生组，这样教师可以为不同的学生组建立个性化学习系统，提供自适应内容，促进有效的小组学习。可以根据学生的技能利用聚类技术和贝叶斯网络混合技术进行分组，还可以根据学生考试分数和在线学习记录，利用 K-means 算法进行分组。

四、未来研究趋势

数据挖掘技术在教育教学领域的应用前景已得到了国内外研究者的肯定，但其发展速度依旧缓慢，想要达到数据挖掘在其他领域的应用成就，还有许多工作需要做。

（一）开发专门的 EDM 工具

由于传统的数据挖掘技术过于复杂，应用难度也较大，且现有的一些广泛应用的数据挖掘工具并不能直接适用于教育领域，所以需要针对教育领域的特点，开发专门的、便于教育工作者使用的 EDM 工具。

(二) 在教育环境中集成 EDM 工具

数据挖掘工具须集成到数字化学习环境中。从数据预处理到结果解释,所有的教育数据挖掘任务都必须在一个应用程序中进行。这样,教育数据挖掘的反馈结果可以直接应用于数字化学习环境,EDM 工具才能被教育工作者广泛使用。

(三) 将 EDM 数据及模型标准化

教育数据资源的收集还未形成统一的数据标准,教育系统中的教育数据挖掘工具还未普适化。为了更规范地进行预处理、知识发现等任务,标准化、规范化教育数据以及数据输入、输出模型是必要的。

(四) 研究特定的 EDM 技巧

EDM 研究不仅涉及教育学、心理学,还需要计算

机科学及统计学的支持。为了尽可能准确地描述 EDM 研究对象的属性,可通过整合教育领域知识、数据挖掘算法,寻求特定的 EDM 方法以获得精准有效的量化数据,从而促进教育数据管理、教学设计和教学决策,这是未来 EDM 研究的重要课题。

五、结语

EDM 是一个融合了计算机科学、心理学、统计学等领域的跨学科领域,它利用统计、机器学习、数据挖掘等技术分析教育数据并为改善教育现状做出决策。EDM 领域近年来发展较快但并未成熟,想要得到广泛应用,不仅需要研究者的努力,还需要教师、学生等各个不同群体的认可和支持。5

参考文献

- [1] Koedinger, D' Mello, McLaughlin, et al. Data mining and education[J]. Wiley Interdisciplinary Reviews: Cognitive Science, 2015, 6: 333-353.
- [2] 李婷, 傅钢善. 国内外教育数据挖掘研究现状及趋势分析[J]. 现代教育技术, 2010, 20 (10): 21-25.
- [3] 顾小清, 林仕丽, 袁海军. 教育数据 30 年: 从 CMI 到 DDDM[J]. 电化教育研究, 2010, (9): 55-63.
- [4] 魏顺平. 教育数据挖掘: 现状与趋势[C]. 信息化、工业化融合与服务创新——第十三届计算机模拟与信息技术学术会议论文集. 2011.
- [5] 葛道凯. E-Learning 数据挖掘: 模式与应用[J]. 中国高教研究, 2012, (3): 8-14.
- [6] 张驰, 陈刚, 王敏娟, 等. 移动学习中使用 EM 算法的学生聚类分析[J]. 中国远程教育, 2009, (5): 68-71.
- [7][8] Romero, Ventura. Educational Data Mining: A Review of the State of the Art[J]. IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, 2010, 40 (6): 601-618.

(责任编辑 孙震华)

Research Status and Trend of Educational Data Mining

DING Bo, SUN Li

(School of Humanities, Jiangnan University, Wuxi, Jiangsu, China 214000)

Abstract: This paper has summarized the rise, development, and future research trends, etc. of Educational Data Mining (EDM), the purpose of which is to provide real-time and valuable reference for all kinds of groups' research and practice of educational data mining. Firstly, this paper describes the background of EDM, and then focuses on the three elements of influencing EDM application, namely, the goal of educational data mining, educational data sources, and educational data mining process, and finally analyzes the EDM application and future research trends.

Key words: Educational Data Mining; influence factors; application; research trend